Review TRENDS in Genetics Vol.21 No.4 April 2005

SL trans-splicing: easy come or easy go?

Kenneth E.M. Hastings

Montreal Neurological Institute and Department of Biology, McGill University, 3801 University St, Montreal, Quebec, Canada H3A 2B4

Is spliced-leader (SL) trans-splicing an ancestral eukaryotic characteristic that has been lost in multiple lineages, or did it arise independently in the various phyla in which it occurs? Recent studies have discovered SL trans-splicing in new metazoan phyla, including the chordates. Its discovery in chordates identifies, for the first time, a phylum that clearly contains both trans-splicing and non-trans-splicing major groups, and defines a limited and well-understood field in which to study the evolutionary dynamics of SL trans-splicing. In this article, I summarize the evolutionarily relevant aspects of SL trans-splicing and consider the interplay among SL trans-splicing, pre-mRNA splice-signal syntax and evolutionary genomics.

Introduction

Spliced-leader (SL) trans-splicing is the spliceosomal transfer of a short RNA sequence, the spliced leader, from the 5′-end of a specialized non-mRNA molecule, the SL RNA (see Glossary), to unpaired splice-acceptor sites on pre-mRNA molecules. As a result, diverse mRNAs – ranging from a minority to 100% of the mRNA population in different organisms – acquire a common 5′-sequence. SL trans-splicing has a patchy phylogenetic distribution and enigmatic evolutionary origins [1]. The evolutionary relevance of SL trans-splicing has recently been broadened by a significant expansion of the known phylogenetic range to include the chordates [2,3]. Here, I provide an updated overview of evolutionary aspects of SL trans-splicing, some of which were briefly reviewed by Nilsen in this journal in 2001 [1], and consider the possible genomic causes and consequences of the gain or loss of SL trans-splicing in an organismal lineage. See Refs [4,5] for a review of the mechanism of trans-splicing and additional reviews considering the origin, function and phylogenetic distribution of SL trans-splicing [1,6,7] and its role in polycistronic transcription [8].

Eukaryotic pre-mRNA cis- and trans-splicing

Spliceosomal pre-mRNA cis-splicing (i.e. intron removal) is an ancient eukaryotic process [9] that is preserved in all or most lineages, albeit much reduced in some groups (e.g. kinetoplastids [10] and yeast [11]). In addition to cis-splicing, the spliceosome is capable of trans-splicing (i.e. joining two separate RNA molecules to form a chimeric molecule). Two forms of spliceosomal trans-splicing occur naturally: (i) pre-mRNA donor trans-splicing, in which a splice-donor site on one pre-mRNA molecule is joined to an acceptor site on another that is, in most cases, transcribed from the same gene or genomic region [12,13]; and (ii) SL trans-splicing.

In SL trans-splicing the donor is a specialized non-mRNA, the SL RNA, whose only known function is to donate a short (~15–50 nt) leader sequence that is transferred to pre-mRNA splice-acceptor sites, and which becomes the 5′-end of the mature mRNA. Current knowledge of the phylogenetic distribution of SL trans-splicing has emerged in two phases. SL trans-splicing was first discovered ~20 years ago in kinetoplastids in the protist phylum Euglenozoa (reviewed in Ref. [14]) and, over the next several years, in two protostome bilaterian metazoans, the nematode *Caenorhabditis elegans* [15] and the flatworm (Platyhelminthes) *Schistosoma mansoni* [16], and also in the protist *Euglena gracilis* (related to kinetoplastids) [17]. In the decade following these initial findings, there were no reports of SL trans-splicing in additional phyla. However, during that time SL trans-splicing was discovered in additional species within the Euglenozoa, nematodes and flatworms, and extensive studies, particularly focused on kinetoplastids and

Glossary

Appendicularian: (also called larvacean) a member of a planktonic tunicate class in which the larval tail is maintained into adulthood.
Ascidian: (also called sea squirt) a member of a sessile tunicate class featuring a dramatic metamorphosis in which the larval tail is lost.
Kinetoplastid: trypanosome; a protist of the phylum Euglenozoa.
Outron: the 5′-segment of a trans-spliced pre-mRNA upstream of the trans-splice-acceptor site.
SL RNA: spliced-leader donor RNA.
Sm proteins: a class of related proteins that bind to a U-rich RNA motif and are found in snRNPs.
snRNP: small ribonucleoprotein particle consisting of snRNA complexed with specific proteins.
Splice-acceptor site: the intron–exon junction at the 3′-end of an intron.
Splice-donor site: the exon–intron junction at the 3′-end of an exon.
TMG: trimethyl guanosine (m_3 2,2,7G), found in the 5′-cap structure of Sm-binding U snRNAs and in metazoan SL RNA.
Tunicate: (also called urochordate) a member of a protochordate subphylum of the chordates consisting of filter-feeding marine organisms having a leathery outer covering or tunic and a larval stage marked by a tail including dorsal nerve cord, notochord and flanking muscle cells.
U snRNAs: a U-rich class of small nuclear RNAs, including spliceosomal snRNAs.
Unpaired splice acceptor: a splice-acceptor site for which there is no corresponding upstream splice-donor site.

Corresponding author: Hastings, K.E.M. (ken.hastings@mcgill.ca).

 0168-9525/$ - see front matter doi:10.1016/j.tig.2005.02.005

nematodes, revealed functional and mechanistic features of SL trans-splicing, as briefly summarized in the following paragraphs.

SL RNAs

Each SL trans-splicing organism has one, or a few, SL RNA species. SL RNAs are transcribed by RNA polymerase II (Pol II) from intronless genes that reside on short (from less than one kilobase to several kilobases) tandemly repeated DNA sequences that, in some organisms, also include 5S rRNA genes [4,18,19]. The presence or absence of the 5S RNA gene in the SL repeat is evolutionarily highly plastic and is not an informative long-range phylogenetic signal [18]. SL RNAs are short RNAs (~45–140 nt) that contain a splice-donor site but no acceptor site, and a hypermodified 5′-cap structure containing 2,2,7-trimethylG (TMG) (in metazoa) [4] or monomethyl m7G in a hypermethylated 'cap4' structure (in kinetoplastids) [20]. The splice-donor site functionally divides the molecule into two segments (Figure 1). During splicing, the 5′-segment (i.e. the leader sequence) behaves like the first exon in a conventionally-expressed gene, and the 3′-segment behaves like the 5′-part of a conventional intron.

Although they have no significant sequence identity, SL RNAs have a striking overall similarity to Sm-class U-rich small nuclear RNAs (snRNAs) (i.e. U1, U2, U4 and U5), which are present in spliceosomal small ribonucleoproteins (snRNPs) and participate in the splicing mechanism (reviewed in Refs [1,4]). Shared features include a small size, TMG cap and a U-rich Sm-protein-binding site (in the SL RNA intron-like moiety). SL RNAs are associated with Sm proteins [21,22] and specific non-Sm proteins that interact *in vivo* with other splicing components [23] in snRNPs. Whether the U snRNP-like nature of the SL snRNA and its unique base-pairing interaction with the catalytically central U6 snRNA of nematodes [4] are simply a means for invading the spliceosome to gain effective access to pre-mRNA splice-acceptor sites, or whether the SL snRNP also makes a non-substrate contribution to the spliceosomal catalytic mechanism, is unknown. SL RNAs from different metazoan phyla have diverse nucleotide sequences. Apart from the presence of the unpaired splice-donor site, which is often involved in a predicted intramolecular duplex secondary structure [7], they do not share features beyond those also shared with the U snRNAs.

Functions of SL trans-splicing

SL trans-splicing has several distinct roles in mRNA function, including: (i) providing a 5′-cap structure for protein-coding RNAs transcribed by the rRNA polymerase, Pol I (uniquely in kinetoplastids [24,25]; Figure 1a); (ii) resolving polycistronic Pol II transcripts into individual capped, monocistronic mRNAs by SL trans-splicing to unpaired acceptor sites that are located upstream of each open reading frame (ORF) (in kinetoplastids [19], nematodes [26] and flatworms [27]; Figure 1b); and (iii) enhancing mRNA translational efficiency through the hypermodified cap structure and/or leader sequence (in kinetoplastids [28] and nematodes [29,30]; Figure 1d).

An additional but less well-understood role for SL trans-splicing is that of trimming and sanitizing the

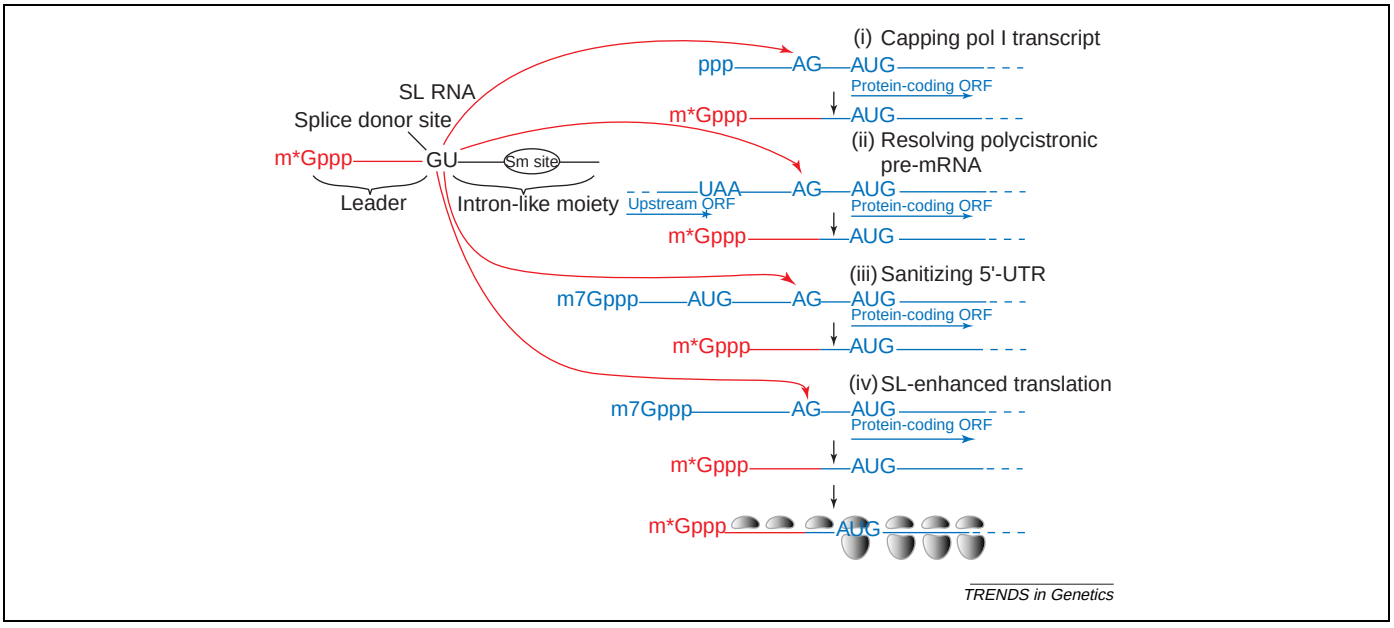

Figure 1. An overview of the biological functions of SL trans-splicing. On the left, the SL RNA is shown with its hypermodified cap structure and leader segment (red), and splice-donor site (indicated by GU) and intron-like moiety with a binding site for Sm proteins (black). The SL RNA cap depiction, m*Gppp, represents either the TMG-containing cap of metazoan SL RNA or the m7G-containing, but otherwise hypermodified, cap4 structure found in kinetoplastid SL RNA. Red arrows indicate trans-splicing events that transfer the SL RNA leader segment to unpaired splice-acceptor sites (indicated by AG) in several types of target pre-mRNA (on which the AUG translation start codons are shown). These trans-splicing events have the following effects: **(i)** uncapped and therefore untranslatable RNA polymerase I transcripts receive a cap structure and become translatable; **(ii)** otherwise untranslatable internal cistrons in polycistronic pre-mRNAs are resolved as separate, capped and translatable mRNAs; **(iii)** pre-mRNAs that are unstable or untranslatable due to problematic elements present in the 5′-untranslated region (5′-UTR) (in this example, an additional upstream AUG codon that would interfere with efficient translation of the proper ORF) are sanitized by removal of the problematic elements along with the discarded 5′-segment of the pre-mRNA (the outron); and **(iv)** in some organisms, efficient translation (indicated by binding of ribosomal small subunits to the mRNA 5′-end, assembly of complete ribosome with large and small subunits at the AUG initiation codon and passage of ribosomes down the mRNA during translation) requires the hypermodified cap structure and/or sequence features of the leader.

5′-untranslated regions of pre-mRNA molecules (Figure 1c) [7]. The 5′-part of a monocistronic pre-mRNA upstream of the trans-splice acceptor site, the outron [31], is discarded during trans-splicing; therefore, it is immaterial whether this sequence includes elements that might compromise mRNA transport, translation or stability (e.g. out-of-frame upstream AUG codons). By contrast, in non-trans-spliced genes it is essential that the initial segment of the primary transcript (i.e. the first exon) is free of such elements because the entire exon is maintained in the mature mRNA. By trimming and sanitizing the 5′-untranslated region, SL trans-splicing probably facilitates the evolution of new transcription-initiation sites [31] and might permit genes to contain a wider variety of potentially useful sequence components near the 5′-end of the primary transcript (e.g. transcriptional regulatory elements).

On the basis of limited data, SL trans-splicing does not appear to be specifically associated with particular tissues, developmental stages, sexes [2,3,32] or with functionally distinct gene classes [7], although it has been noted that the minority fraction of *Caenorhabditis* trans-spliced mRNAs that are derived from polycistronic operon transcripts largely encode ubiquitously expressed housekeeping proteins [8]. Housekeeping genes are especially compatible with operon organization because they presumably do not require independent transcriptional control [8]. In different phyla, a varying fraction of the mRNA population is trans-spliced. In the protist kinetoplastids, apparently, all mRNA molecules are SL trans-spliced because most or all genes are expressed as SL-resolved operons [19]. In metazoa, the trans-spliced fraction ranges from a large majority, 70–90% in the nematodes *Caenorhabditis* and *Ascaris* [6,29], to an imprecisely known minority in the flatworm *Schistosoma* [32]. Operon resolution cannot account for the majority of SL trans-splicing in nematodes, where only ~15% of the genome is comprised of operons [8], but we do not yet know whether the same is true for flatworms [or other metazoan phyla (discussed in the following sections)].

Pre-mRNA splice-signal syntax

Cis- and trans-splicing coexist in SL trans-splicing organisms. Whether cis- or trans-splicing occurs at any particular splice-acceptor site is primarily determined by pre-mRNA splice-signal syntax. Proceeding from the 5′-cap in a conventionally expressed non-trans-spliced pre-mRNA, the first splice signal encountered is the donor site at the junction of exon one and intron one. This is donor-first syntax. Acceptor-first syntax occurs when a splice-acceptor site exists upstream of the 5′-most donor site (Figure 2a). Only pre-mRNAs that have an acceptor-first syntax undergo 5′-end SL trans-splicing (similarly, only unpaired internal splice-acceptor sites that also lack an upstream donor are targeted in SL-resolved polycistronic transcripts). By contrast, donor-first pre-mRNAs are not targets for trans-splicing; they preferentially undergo cis-splicing (i.e. removal of the introns defined by paired donor and acceptor sites). Apparent 'illegitimate' partial trans-splicing at paired cis-splice-acceptor sites at the ends of rather long introns has been reported [10,16,33,34]; however, in at least one case, this appears to reflect trans-splicing of alternative transcripts initiated from additional promoters within the intron [33] and having legitimate acceptor-first syntax. The spliceosome clearly has a strong preference for a cis-linked donor, if one is available. The mechanistic basis for this preference is unknown.

In the nematodes *Caenorhabditis* and *Ascaris*, experimental conversion of donor-first to acceptor-first syntax can be sufficient to direct SL trans-splicing of otherwise non-trans-spliced pre-mRNAs. However, the trans-splicing efficiency can be markedly affected by downstream elements that promote exon definition and recognition of splice-acceptor sites [35–39] and by the length and base composition of the outron [37]. Apart from syntactic context, *Caenorhabditis* and *Ascaris* trans-splice-acceptor sites are functionally similar to cis-splice-acceptor sites. By contrast, in the filiarial nematode *Brugia,* trans-splicing requires a specific 7-nt motif, located in the first downstream intron, which is not required for cis-splicing [40]. Given that *Brugia* resides within the *Caenorhabditis* and *Ascaris* clade [41], this might be a lineage-specific innovation – one whose biological significance is unclear.

The interplay between SL trans-splicing and pre-mRNA splice-signal syntax is likely to be an important factor in evolutionary genomics.

SL trans-splicing in the chordates

Despite accelerating molecular-genetic research since the initial phase of trans-splicing discovery, no examples of SL trans-splicing were found in several intensively studied groups (e.g. vertebrates, arthropods, plants and fungi). Thus, unlike cis-splicing, SL trans-splicing does not appear to be a universal eukaryotic mechanism, and for some time it seemed that it might be limited to those groups in which it was initially discovered. However, in recent years, there has been a second phase of trans-splicing discovery involving two additional metazoan phyla, the diploblastic (pre-bilaterian) cnidarians [42] and the chordates [2,3]. These discoveries had a major impact on the known distribution of SL trans-splicing (e.g. the chordate discoveries were the first in the deuterostome division of the metazoa) (Figure 3). An additional major significance of the chordate discoveries is that, for the first time, they place trans-splicing within a phylum known to contain a major group – the vertebrates – that does not perform trans-splicing. This has significant implications for evolutionary studies (discussed in the next section).

In chordates, the initial identification of SL trans-splicing was in the ascidian tunicate *Ciona intestinalis* [2], following in-depth studies of the gene encoding troponin I. The full extent of SL trans-splicing in *Ciona* has not been established, but at least seven different mRNAs, and possibly many more, contain a common trans-spliced leader [2]. More recently, SL trans-splicing was also discovered in an additional tunicate class, the Appendicularia (Larvacea), based on global EST and genome sequencing results. In the appendicularian *Oikopleura dioica* 12–24% of mRNAs are trans-spliced [3], in many cases reflecting SL-resolved operon transcripts. Polycistronic operons have not yet

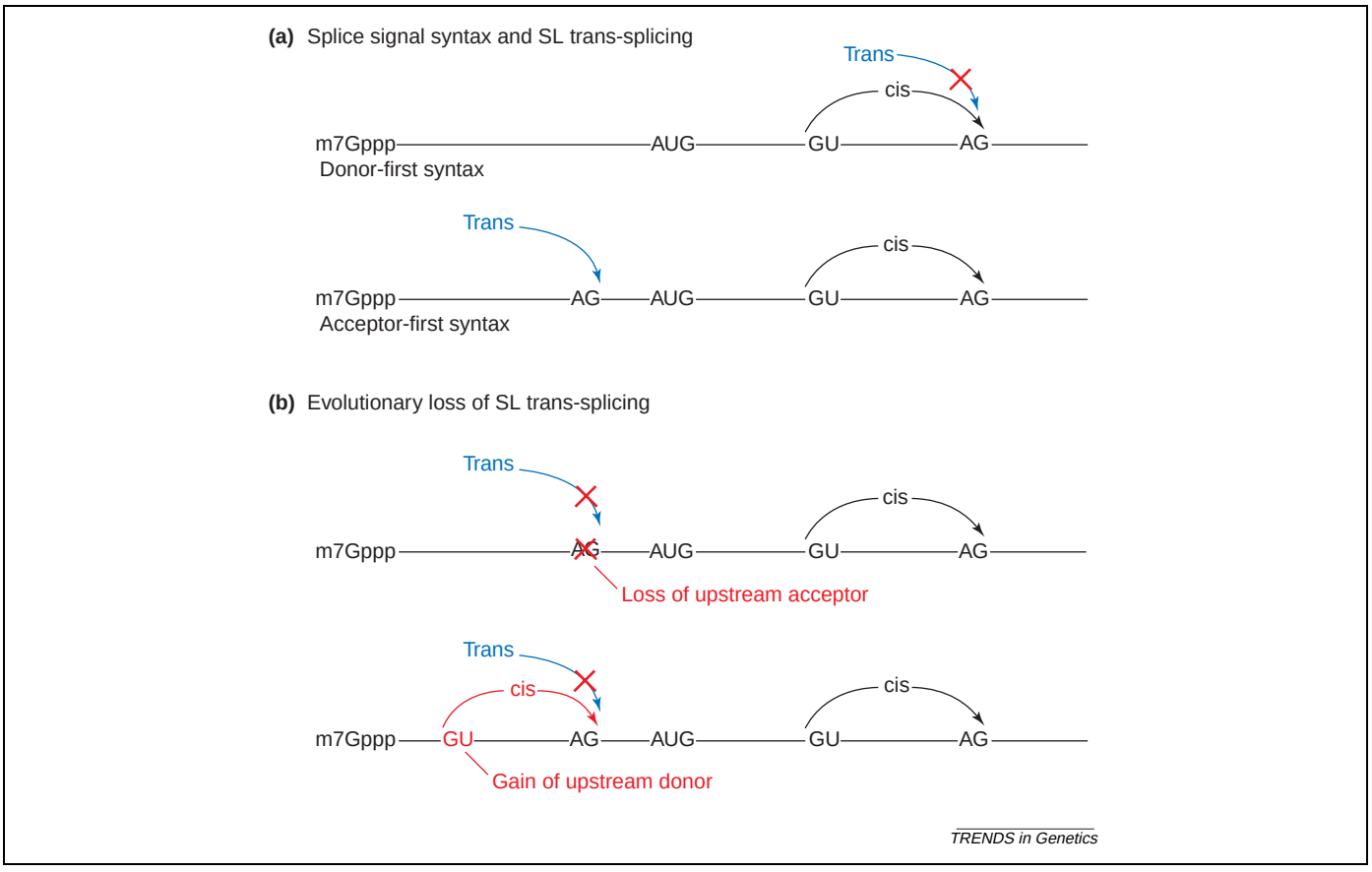

Figure 2. SL trans-splicing and pre-mRNA splice-signal syntax. The figure depicts a series of unspliced pre-mRNAs, and the splicing reactions they can undergo. The standard pre-mRNA 5′-cap structure m7Gppp, AUG translation start codons and splice acceptor (AG) and donor (GU) sites are shown. Each pre-mRNA shown contains an intron downstream of the AUG start codon. **(a)** Examples of donor-first and acceptor-first splice-signal syntax, in which the 5′-most splice site is either a donor or an acceptor, respectively. In the case of donor-first syntax, trans-splicing (blue arrow) cannot compete with conventional cis-splicing (black arrow). In the example of acceptor-first syntax, trans-splicing can occur at the upstream unpaired splice-acceptor site because there is no cis-splicing competition. **(b)** Evolutionary loss of trans-splicing by two kinds of mutations that change acceptor-first syntax to donor-first syntax: loss of the trans-splice-acceptor site, and gain of a new upstream splice-donor site which, by permitting cis-splicing, would preclude trans-splicing.

been reported in ascidians, although, based on genomic organization, a *Ciona* trans-spliced mRNA encoding a proteasome subunit appears to derive from the upstream member of a bicistronic operon; the remaining six known *Ciona* trans-spliced mRNAs appear to derive from monocistronic genes (K.E.M. Hastings unpublished data). The tunicate SL RNAs contain a TMG 5′-cap [3] (K.E.M. Hastings unpublished data), are associated with Sm proteins [3] and are encoded by tandemly-repeated genomic DNA sequences. In *Oikopleura,* the SL RNA gene repeat includes the 5S RNA gene [3], but this is not the case in *Ciona* (K.E.M. Hastings, unpublished data).

Evolution of SL trans-splicing

Within the euglenozoa, nematodes, flatworms and tunicates, SL trans-splicing is phylogenetically widespread, suggesting that it is an ancient feature in each of these groups (Box 1). (The extent of SL trans-splicing in the cnidarians is presently unknown.) This, coupled with the wide distribution of SL trans-splicing phyla in the eukaryotic evolutionary tree (Figure 3), is consistent with the view that SL trans-splicing could have been present in the ancestral eukaryote. According to this unique-origin hypothesis, those intensively-studied eukaryotes in which it has not yet been found represent cases where ancestral SL trans-splicing has been lost.

Although the unique-origin hypothesis is plausible, the counterhypothesis – that SL trans-splicing arose independently in the multiple eukaryotic phyla where it has been found – cannot be eliminated. SL RNAs from different phyla do not show significant direct sequence similarity and clear orthologues of nematode SL snRNP-specific proteins [23] and TMG-binding initiation factor 4E isoforms [43] have not been found in other organisms [3,23,42]. In addition, homologous genes that are trans-spliced in one phylum are not necessarily trans-spliced in another [7]. However, these phylum-specific features could simply reflect extensive divergence and no compelling evolutionary case has yet been made for the *de novo* appearance of SL-trans-splicing in any particular group. Further studies of the chordates, which include both trans-splicing and non-trans-splicing groups, will probably clarify the relative roles of gain and loss within this phylum. However, to understand the evolutionary relationships among SL trans-splicing processes across the eukaryotes might require detailed comparative mechanistic knowledge for each phylum.

Notwithstanding uncertain relationships at the highest taxonomic levels, there is good evidence for a common

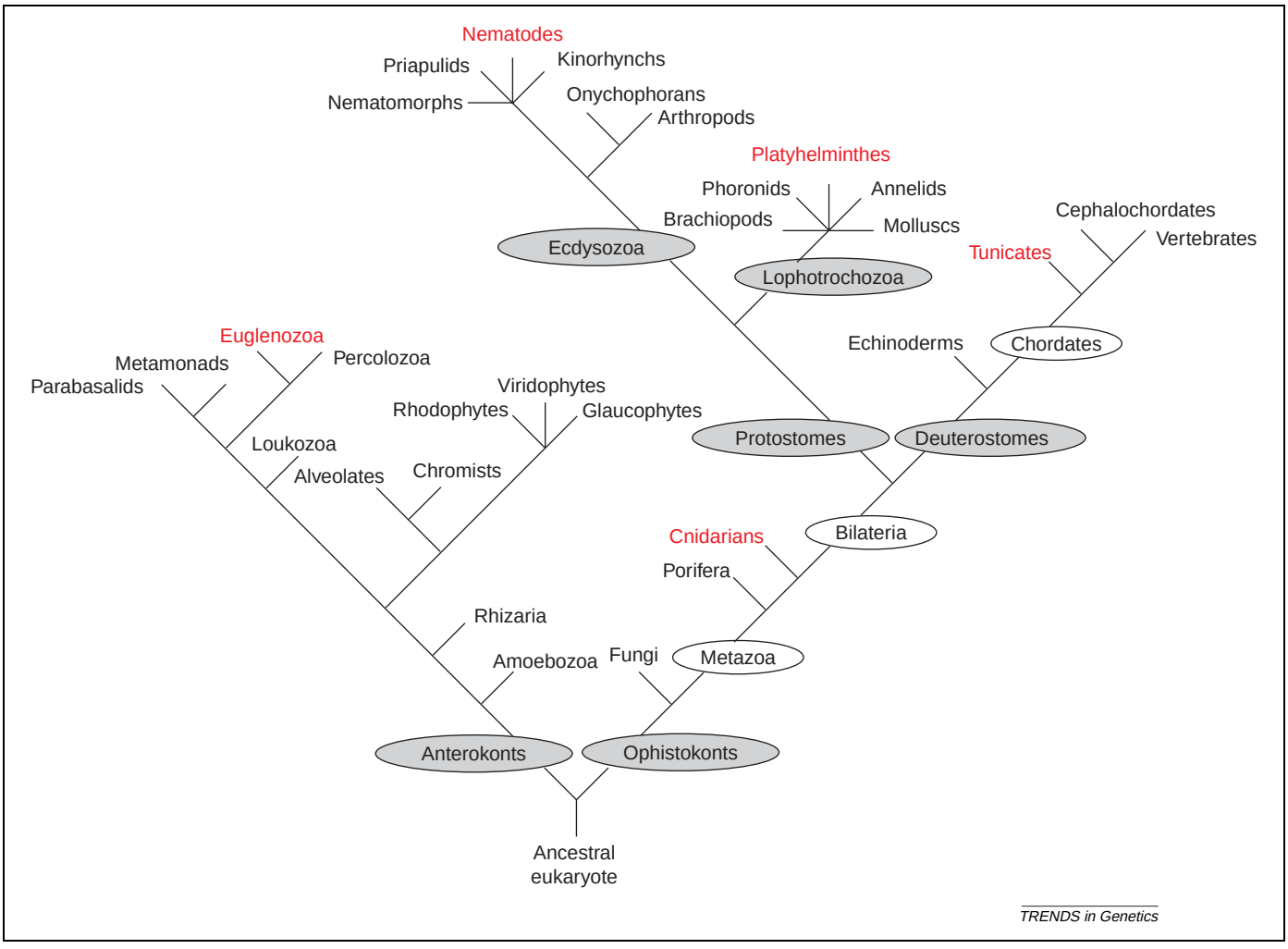

Figure 3. Phylogenetic distribution of known examples of SL trans-splicing in the eukaryotes. The tree shown is a global eukaryotic tree as developed by Cavalier-Smith [63] with within-metazoa details as in Erwin and Davidson [64]. Groups in which SL-trans-splicing has been observed are shown in red. Note that examples of trans-splicing are broadly distributed throughout the eukaryotic evolutionary tree.

origin for trans-splicing among the species within each major group (Box 1). This makes it possible to consider whether the trans-splicing status of individual genes is evolutionarily stable or highly plastic. Most (96%) of the 900 known SL-resolved operons of *Caenorhabditis elegans* are also present as operons in *Caenorhabditis briggsae* [divergence time is 50–100 million years (Myr)] [8,44] and, although there are exceptions [45], at least several operons seem to be widely conserved within the nematodes [8]. Flatworm enolase, in an SL-resolved operon, is trans-spliced in a wide variety of flatworm species, including the distantly related *Schistosoma* and *Stylochus* [46]. Thus, operon-associated trans-splicing can be evolutionarily stable. There is little comparative data concerning trans-spliced monocistronic genes; however, one such gene, encoding troponin I, is trans-spliced in the distantly-related ascidians *Ciona* [2] and *Halocynthia* [47]. Comprehensive genome-wide studies of related species are needed to fully characterize the evolutionary plasticity of trans-splicing. Such comparative genomics data are likely to be available soon for the genus *Caenorhabditis* (species *elegans* and *briggsae*) [44,48] and the genus *Ciona* (species *intestinalis* and *savignyi*) ([49] and http://www.broad.mit.edu/annotation/ciona).

Evolutionary gain or loss of SL trans-splicing

The phylogenetic distribution of organisms known to perform SL trans-splicing implies that it has either been independently invented, or independently lost, multiple times [42]. How could SL trans-splicing be gained or lost and what would the impact be?

Gain

SL RNAs could arise *de novo* from an Sm-binding U snRNA (by evolution of a splice-donor site) or from a fragment of a protein-coding gene including the promoter, the first exon and part of the first intron (by evolution of an Sm-binding site in the intron). The *de novo* appearance of an efficient SL RNA could be deleterious if essential protein-coding sequences contained cryptic unpaired splice-acceptor sites, because these would immediately become targets for ORF-disrupting SL trans-splicing events. However, if the initial emergence of the SL RNA is tolerated (e.g. if no key genes have cryptic acceptors or if trans splicing is relatively inefficient at first), then subsequent evolution could exploit novel possibilities associated with SL trans-splicing. For example, SL trans-splicing permits the consolidation of individual genes into polycistronic operons, which is a more efficient

Box 1. SL trans-splicing: an ancient feature of euglenozoa, flatworms, nematodes and tunicates

Euglenozoa

In the protists, trans-splicing has been identified in two major classes within the Phylum Euglenozoa, kinetoplastids and euglenoids. Because the kinetoplastid–euglenoid divergence is an ancient one [57], SL trans-splicing is likely to be an ancestral feature in this phylum. There is little or no sequence similarity of the euglenoid and kinetoplastid spliced leaders, presumably reflecting extensive sequence divergence. Divergence is also evident within the euglenoids; the 26–27-nt euglenoid leader sequence is nearly identical in closely related species [58]; however, only several nucleotides at the 5′ and 3′ ends appear to be conserved across distantly-related genera [59]. There is significant, but lesser, divergence among kinetoplastid spliced leaders, where 23 residues of the 39–41-nt leader sequence are conserved across many genera [60].

Platyhelminthes (flatworms)

The flatworms are divided into four classes, Turbellaria (planarians), Trematoda (flukes), Cestoda (tapeworms) and Monogenea (ectoparasitic flukes) and, to date, SL trans-splicing has been found in the first three [33,46]. This wide distribution, coupled with the likelihood that the turbellaria are the most primitive class, is consistent with ancestral SL trans-splicing in this phylum. The ~36-nt spliced leaders of trematodes and cestodes are ~50% identical [33]. Although the turbellarian spliced leader is longer (51 nt) and more divergent, ~36% identical in aligned regions [46], there is, nonetheless, sufficient similarity to clearly indicate descent of all flatworm SL RNAs from a common ancestor.

Nematodes

SL trans-splicing is known to occur in many species of nematode, of which the best-studied is *Caenorhabditis elegans* (Class Secernentea). Various nematode phylogenies agree that the trichocephalid *Trichinella spiralis* is separated from the Secernentea by the oldest known divergence within the phylum [41]. SL trans-splicing is thought to occur in *Trichinella spiralis* [6] (although a full report has yet to appear), which would indicate that SL trans-splicing is an ancestral nematode characteristic. Within the Secernentea, SL trans-splicing is widespread, including the ascaridid *Ascaris*, which is separated from *Caenorhabditis* by the oldest known divergence within the group [41]. Moreover, the sequence of the 22-nt SL1 leader is identical throughout the Secernentea, although the intron-like remainder of the SL RNA sequence varies [6]. In *Caenorhabditis* and several related genera (in Clade V in Ref. [41]), but in not more distantly-related groups, an additional specialized operon-resolving SL RNA, SL2, resembling the major SL RNA, SL1, is present [6,8]. SL2 is specifically recruited to internal cistrons of polycistronic transcripts by association with proteins involved in mRNA 3′-end formation at the upstream cistrons [61]. In other nematodes, and in other phyla, one SL RNA is used both for monocistronic and polycistronic trans-splicing.

Tunicates

Chordate SL trans-splicing was discovered in the ascidian *Ciona intestinalis* [2], and recent evidence [47] strongly suggests that SL trans-splicing also occurs in *Halocynthia roretzi*, a pyurid ascidian separated from *Ciona* by the oldest known divergence among ascidians [56,62]. The *Halocynthia* putative leader sequence is 24 nt rather than 16 nt as in *Ciona*, but except for two short indels these two sequences are very similar. The appendicularian *Oikopleura dioica* perfroms SL trans-splicing with a 40-nt leader whose sequence is differs considerably from that of *Ciona* [3]. The appendicularian–ascidian separation is the oldest known divergence within the tunicates [56,62], suggesting that SL trans-splicing was an ancestral tunicate character.

use of genomic DNA [3,8]. Thus, the presence of trans-splicing is a facilitatory factor for evolutionary genome compaction and could have contributed, for example, to the marked genome compaction of appendicularians [3].

Loss

SL trans-splicing could be lost instantaneously through deletion of the genes encoding the SL RNA itself or an essential SL snRNP-specific protein component [23]. Such instantaneous global loss of SL trans-splicing could be deleterious, as shown by the lethality of spliced-leader 1 (SL1) gene deletion in *Caenorhabditis* [50]. Alternatively, and more likely, SL trans-splicing could be lost gradually over evolutionary time on a gene-by-gene basis, by the mutational conversion of operons to monocistronic genes and of trans-spliced monocistronic genes from acceptor-first syntax to donor-first syntax. Donor-first syntax could arise either by mutational loss of the unpaired trans-splice-acceptor site or by the emergence of a new splice-donor site upstream of the unpaired trans-splice-acceptor site (Figure 2). By consigning much of the erstwhile outron to intron status, the second type of mutation described could preserve much of the trimming and sanitizing function of ancestral trans-splicing. Operon breakup would probably be favored in lineages undergoing evolutionary genome expansion by DNA insertion and/or duplication via: (i) insertion of new DNA elements (possibly containing functional promoters) between cistrons; and (ii) loss of complementary sets of cistrons from initially identical duplicates of ancestral operons, consistent with the duplication-degeneration-complementation (DDC) model for preservation of duplicate loci by subfunctionalization [51].

The discovery of SL trans-splicing in the chordates raises the possibility that vertebrates might have descended from a trans-splicing ancestor [2]. If so, the significant genome expansion that accompanied vertebrate evolution [52,53] probably contributed to conditions that favored operon breakup and the loss or reduction of ancestral SL trans-splicing. However, not all non-SL-trans-splicing lineages are associated with evolutionary genome expansion (e.g. arthropods and fungi); therefore, other contributing genomic evolutionary trends would have to be envisaged. Moreover, operon breakup would not account for the loss of trans-splicing from monocistronic genes. One can readily conceive of global pressures favoring the gene-by-gene transformation of acceptor-first into donor-first syntax [e.g. pressure to have a splice-donor site (or intron) close to the transcription start site]. Given the tight integration of the splicing machinery into upstream and downstream processes in mRNA production [54] (e.g. the stimulation of transcriptional elongation by intron-recruited splicing factors [55]), evolutionary change in any aspect of the gene-expression machinery could introduce novel pressures on the distribution of pre-mRNA splicing signals, and thus lead to evolutionary change in the population of genes undergoing trans-splicing.

Future studies

Our understanding of the evolution of SL trans-splicing would be greatly advanced by a clear-cut example of either

gain or loss. The chordates are uniquely well-suited to address this issue because this is the only phylum currently known to have both trans-splicing and non-trans-splicing groups. A diligent search for SL-trans-splicing in other members of the phylum, and in related deuterostome phyla, might reveal a distribution that strongly favors either *de novo* emergence in the tunicates or loss in the vertebrates. For example, the finding of SL trans-splicing in the cephalochordates, a protochordate group more closely related to vertebrates than are tunicates [56], would strongly suggest loss in the vertebrate lineage. However, the absence of SL trans-splicing from cephalochordates and other deuterostome phyla (e.g. hemichordates and echinoderms) would strongly suggest the *de novo* emergence of SL trans-splicing in the tunicates.

The 'second phase' of trans-splicing discovery has broadened the known distribution and raised the possibility of even further discoveries. The most recent example, the appendicularians, is the first discovery of SL trans-splicing to arise from a genome project [3], as opposed to detailed investigation of individual genes. The implementation of genome projects for an increasing variety of organisms is likely to lead to further expansion of the known phylogenetic range of SL trans-splicing. Any such discovery in additional phyla would, by an accumulating advantage of parsimony, favor the unique-origin hypothesis [2]. However, definitively establishing whether SL trans-splicing is an orthologous process in all of the eukaryotic phyla in which it occurs is likely to require comparative analysis of detailed molecular mechanisms. Mechanistically, nematode trans-splicing is the best characterized and this pioneering work has already suggested several specific lines of inquiry. For example, do SL RNAs in groups other than nematodes have a similar interaction with U6 snRNA? Do SL snRNP-specific proteins exist in other phyla than nematodes, and what is their evolutionary relationship to the nematode SL snRNP-specific proteins? Clear answers to these, and other, questions, that will arise as our understanding of the nematode trans-splicing mechanism improves, should be insightful.

Acknowledgements

I thank E.M. Thompson and P. Ganot for sharing results in advance of publication. This research is supported by the Natural Sciences and Engineering Research Council of Canada.

References

1 Nilsen, T.W. (2001) Evolutionary origin of SL-addition trans-splicing: still an enigma. *Trends Genet.* 17, 678–680
2 Vandenberghe, A.E. *et al*. (2001) mRNA 5′-leader trans-splicing in the chordates. *Genes Dev.* 15, 294–303
3 Ganot, P. *et al*. (2004) Spliced-leader RNA trans splicing in a chordate, *Oikopleura dioica*, with a compact genome. *Mol. Cell. Biol.* 24, 7795–7805
4 Nilsen, T.W. (1993) Trans-splicing of nematode premessenger RNA. *Annu. Rev. Microbiol.* 47, 413–440
5 Nilsen, T.W. (1995) Trans-splicing: an update. *Mol. Biochem. Parasitol.* 73, 1–6
6 Blaxter, M. and Liu, L. (1996) Nematode spliced leaders–ubiquity, evolution and utility. *Int. J. Parasitol.* 26, 1025–1033
7 Davis, R.E. (1996) Spliced leader RNA trans-splicing in metazoa. *Parasitol. Today* 12, 33–40
8 Blumenthal, T. and Gleason, K.S. (2003) *Caenorhabditis elegans* operons: form and function. *Nat. Rev. Genet.* 4, 112–120
9 Johnson, P.J. (2002) Spliceosomal introns in a deep-branching eukaryote: the splice of life. *Proc. Natl. Acad. Sci. U. S. A.* 99, 3359–3361
10 Mair, G. *et al*. (2000) A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. *RNA* 6, 163–169
11 Spingola, M. *et al*. (1999) Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* 5, 221–234
12 Maniatis, T. and Tasic, B. (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* 418, 236–243
13 Horiuchi, T. *et al*. (2003) Alternative trans-splicing of constant and variable exons of a *Drosophila* axon guidance gene, lola. *Genes Dev.* 17, 2496–2501
14 Van der Ploeg, L.H. (1986) Discontinuous transcription and splicing in trypanosomes. *Cell* 47, 479–480
15 Krause, M. and Hirsh, D. (1987) A trans-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* 49, 753–761
16 Rajkovic, A. *et al*. (1990) A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni*. *Proc. Natl. Acad. Sci. U. S. A.* 87, 8879–8883
17 Tessier, L.H. *et al*. (1991) Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by trans-splicing in *Euglena*. *EMBO J.* 10, 2621–2625
18 Drouin, G. and de Sa, M.M. (1995) The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. *Mol. Biol. Evol.* 12, 481–493
19 Campbell, D.A. *et al*. (2003) Transcription in kinetoplastid protozoa: why be normal? *Microbes Infect.* 5, 1231–1240
20 Bangs, J.D. *et al*. (1992) Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. *J. Biol. Chem.* 267, 9805–9815
21 Maroney, P.A. *et al*. (1990) The nematode spliced leader RNA participates *in trans*-splicing as an Sm snRNP. *EMBO J.* 9, 3667–3673
22 Palfi, Z. *et al*. (2000) The spliceosomal snRNP core complex of *Trypanosoma brucei*: cloning and functional analysis reveals seven Sm protein constituents. *Proc. Natl. Acad. Sci. U. S. A.* 97, 8967–8972
23 Denker, J.A. *et al*. (2002) New components of the spliced leader RNP required for nematode trans-splicing. *Nature* 417, 667–670
24 Lee, M.G. and Van der Ploeg, L.H. (1997) Transcription of protein-coding genes in trypanosomes by RNA polymerase I. *Annu. Rev. Microbiol.* 51, 463–489
25 Gunzl, A. *et al*. (2003) RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Eukaryot. Cell* 2, 542–551
26 Blumenthal, T. *et al*. (2002) A global analysis of *Caenorhabditis elegans* operons. *Nature* 417, 851–854
27 Davis, R.E. and Hodgson, S. (1997) Gene linkage and steady state RNAs suggest trans-splicing may be associated with a polycistronic transcript in *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* 89, 25–39
28 Zeiner, G.M. *et al*. (2003) The Leishmania tarentolae spliced leader contains determinants for association with polysomes. *J. Biol. Chem.* 278, 38269–38275
29 Maroney, P.A. *et al*. (1995) Most mRNAs in the nematode *Ascaris lumbricoides* are trans-spliced: a role for spliced leader addition in translational efficiency. *RNA* 1, 714–723
30 Lall, S. *et al*. (2004) Contribution of trans-splicing, 5′-leader length, cap-poly(A) synergism, and initiation factors to nematode translation in an ascaris suum embryo cell-free system. *J. Biol. Chem.* 279, 45573–45585
31 Blumenthal, T. (1995) Trans-splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends Genet.* 11, 132–136
32 Davis, R.E. *et al*. (1995) RNA trans-splicing in flatworms. Analysis of trans-spliced mRNAs and genes in the human parasite, *Schistosoma mansoni*. *J. Biol. Chem.* 270, 21813–21819
33 Brehm, K. *et al*. (2002) Characterization of a spliced leader gene and of trans-spliced mRNAs from *Taenia solium*. *Mol. Biochem. Parasitol.* 122, 105–110
34 Spicher, A. *et al*. (1994) Extremely stable transcripts may compensate for the elimination of the gene fert-1 from all *Ascaris lumbricoides* somatic cells. *Dev. Biol.* 164, 72–86

35 Conrad, R. *et al*. (1991) Insertion of part of an intron into the 5′ untranslated region of a *Caenorhabditis elegans* gene converts it into a trans-spliced gene. *Mol. Cell. Biol.* 11, 1921–1926
36 Conrad, R. *et al*. (1993) Conversion of a trans-spliced *C. elegans* gene into a conventional gene by introduction of a splice donor site. *EMBO J.* 12, 1249–1255
37 Conrad, R. *et al*. (1995) SL1 trans-splicing specified by AU-rich synthetic RNA inserted at the 5′ end of *Caenorhabditis elegans* pre-mRNA. *RNA* 1, 164–170
38 Romfo, C.M. *et al*. (2001) 3′ splice site recognition in nematode trans-splicing involves enhancer-dependent recruitment of U2 snRNP. *RNA* 7, 785–792
39 Boukis, L.A. and Bruzik, J.P. (2001) Functional selection of splicing enhancers that stimulate trans-splicing *in vitro*. *RNA* 7, 793–805
40 Higazi, T.B. and Unnasch, T.R. (2004) Intron encoded sequences necessary for trans splicing in transiently transfected *Brugia malayi*. *Mol. Biochem. Parasitol.* 137, 181–184
41 Blaxter, M.L. *et al*. (1998) A molecular evolutionary framework for the phylum Nematoda. *Nature* 392, 71–75
42 Stover, N.A. and Steele, R.E. (2001) Trans-spliced leader addition to mRNAs in a cnidarian. *Proc. Natl. Acad. Sci. U. S. A.* 98, 5693–5698
43 Keiper, B.D. *et al*. (2000) Functional characterization of five eIF4E isoforms in *Caenorhabditis elegans*. *J. Biol. Chem.* 275, 10590–10596
44 Stein, L.D. *et al*. (2003) The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol* 1, E45
45 Lee, K.Z. and Sommer, R.J. (2003) Operon structure and trans-splicing in the nematode *Pristionchus pacificus*. *Mol. Biol. Evol.* 20, 2097–2103
46 Davis, R.E. (1997) Surprising diversity and distribution of spliced leader RNAs in flatworms. *Mol. Biochem. Parasitol.* 87, 29–48
47 Yuasa, H.J. *et al*. (2002) The structural organization of ascidian *Halocynthia roretzi* troponin I genes. *J. Biochem. (Tokyo)* 132, 135–141
48 Hwang, B.J. *et al*. (2004) Genome annotation by high-throughput 5′ RNA end determination. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1650–1655
49 Dehal, P. *et al*. (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298, 2157–2167
50 Ferguson, K.C. *et al*. (1996) The SL1 trans-spliced leader RNA performs an essential embryonic function in *Caenorhabditis elegans* that can also be supplied by SL2 RNA. *Genes Dev.* 10, 1543–1556
51 Force, A. *et al*. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545
52 McLysaght, A. *et al*. (2002) Extensive genomic duplication during early chordate evolution. *Nat. Genet.* 31, 200–204
53 Gu, X. *et al*. (2002) Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. *Nat. Genet.* 31, 205–209
54 Maniatis, T. and Reed, R. (2002) An extensive network of coupling among gene expression machines. *Nature* 416, 499–506
55 Fong, Y.W. and Zhou, Q. (2001) Stimulatory effect of splicing factors on transcriptional elongation. *Nature* 414, 929–933
56 Swalla, B.J. *et al*. (2000) Urochordates are monophyletic within the deuterostomes. *Syst. Biol.* 49, 52–64
57 Preisfeld, A. *et al*. (2001) Phylogenetic position and inter-relationships of the osmotrophic euglenids based on SSU rDNA data, with emphasis on the Rhabdomonadales (Euglenozoa). *Int. J. Syst. Evol. Microbiol.* 51, 751–758
58 Frantz, C. *et al*. (2000) Characterization of trans-splicing in Euglenoids. *Curr. Genet.* 37, 349–355
59 Ebel, C. *et al*. (1999) Trans-splicing and cis-splicing in the colourless Euglenoid, *Entosiphon sulcatum*. *Curr. Genet.* 35, 542–550
60 Sturm, N.R. *et al*. (1995) The mini-exon genes of three Phytomonas isolates that differ in plant tissue tropism. *FEMS Microbiol. Lett.* 130, 177–182
61 Liu, Y. *et al*. (2003) An uncapped RNA suggests a model for *Caenorhabditis elegans* polycistronic pre-mRNA processing. *RNA* 9, 677–687
62 Wada, H. (1998) Evolutionary history of free-swimming and sessile lifestyles in urochordates as deduced from 18S rDNA molecular phylogeny. *Mol. Biol. Evol.* 15, 1189–1194
63 Cavalier-Smith, T. (2002) The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* 52, 297–354
64 Erwin, D.H. and Davidson, E.H. (2002) The last common bilaterian ancestor. *Development* 129, 3021–3032